

Chapter 578

Confidence Intervals for Regression-Based Reference Limits using Assurance Probability

Introduction

This procedure calculates the sample size necessary to achieve a stated assurance probability in exact confidence intervals of regression-based reference limits (percentiles). Using the ratio of the widths of the confidence interval and the reference interval as a relative precision index, optimal sample size can be found using a specific tolerance interval. The procedure is based on the work of Shieh (2017) which shows that the proposed exact confidence intervals are much more precise than the approximate procedures that are commonly used.

Technical Details

These methods used in this procedure come from Shieh (2017). This article provides useful insights into why one should use the new procedure proposed by the author.

Confidence Interval Estimation

Assume that you are planning a study that will generate data commonly analyzed by multiple linear regression model consisting of a response variable Y and K covariate (independent) variables X_1, X_2, \dots, X_K . These variables are measured on N subjects. The regression model is of the form

$$Y_i = \beta_0 + \sum_{k=1}^K X_{ik}\beta_k + e_i$$

Where Y_i is an observed, continuous response variable on the i^{th} subject; X_{i1}, \dots, X_{iK} are observed, continuous covariates; $\beta_0, \beta_1, \dots, \beta_K$ are unknown regression parameters to be estimated; and e_i are *iid* $N(0, \sigma^2)$ random errors for $i = 1, \dots, N$.

Confidence Intervals for Regression-Based Reference Limits using Assurance Probability

Suppose the mean response at a particular value of X , say X_0 , $\mu = E[Y|\{X_{01}, \dots, X_{0K}\}] = \beta_0 + \sum_{k=1}^K X_{0k}\beta_k$. The estimated value of μ is

$$\hat{\mu} = b_0 + \sum_{k=1}^K X_{0k}b_k$$

where the b 's are the regression estimates of the coefficients.

Suppose you want to estimate the 100 p th percentile (reference limit) which is denoted by θ_p , where

$$\theta_p = \mu + z_p\sigma.$$

The minimum variance unbiased estimator is

$$\hat{\theta}_p = \hat{\mu} + cz_p\hat{\sigma}$$

where

$$c = \frac{\sqrt{v/2}\Gamma(v/2)}{\Gamma((v+1)/2)}, \quad v = N - K - 1$$

Here, z_p is the 100 p th percentile of the standard normal distribution $N(0,1)$.

The calculation of the exact two-sided, 100(1 - α)% confidence interval for the reference limit ($\hat{\theta}_{pL}, \hat{\theta}_{pU}$) is done as follows.

$$\begin{aligned} X_i &= (X_{i1}, \dots, X_{iK})' \\ X_0 &= (X_{01}, \dots, X_{0K})' \\ \bar{X} &= \sum_{i=1}^N X_i/N \\ A &= \sum_{i=1}^N (X_i - \bar{X})(X_i - \bar{X})' \\ W &= \frac{1}{N} + \sum_{i=1}^N (X_0 - \bar{X})' A^{-1} (X_0 - \bar{X}) \\ \hat{\sigma}^2 &= \frac{SSE}{v} \\ \hat{\theta}_{pL} &= \hat{\mu} + \sqrt{W} t'_{\frac{\alpha}{2}, v, \frac{z_p}{\sqrt{W}}} \hat{\sigma} \\ \hat{\theta}_{pU} &= \hat{\mu} + \sqrt{W} t'_{1-\frac{\alpha}{2}, v, \frac{z_p}{\sqrt{W}}} \hat{\sigma} \end{aligned}$$

with SSE = error sum of squares from the multiple regression and t' is the percentile of the non-central t distribution with corresponding degrees of freedom and noncentrality parameter. Note that t' is not symmetric so that the confidence interval is not symmetric either.

Sample Size

The sample size calculation is based on the relative precision, Q , which is given by

$$Q = C/R$$

where $C = \hat{\theta}_{pU} - \hat{\theta}_{pL}$ and $R = 2z_p c \hat{\sigma}$. Note that $\hat{\mu}$ and $\hat{\sigma}$ cancel out of the Q ratio.

The sample size is then constructed so that

$$\Pr(Q \leq \omega) \geq 1 - \gamma.$$

Shieh (2017) provides a numerical integration algorithm for calculating this probability, so the sample size can be determined by a simple search.

Example 1 – Finding Sample Size

Suppose a study is planned in which the researcher wishes to construct an exact two-sided 95% confidence interval for the 95th percentile such that the relative precision (Q) is guaranteed to be no wider than 0.1, 0.2, or 0.3 with probability 0.9. The values of λ are 1, 1.2, 1.4 when $K = 1$.

The goal is to determine the necessary sample size.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the procedure window. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Sample Size (N)
Confidence Level (1 - α)	0.95
K (Number of Covariates)	1
1 - γ (Lower Bound of Assurance Probability)	0.90
ω (Upper Bound of the Relative Precision, Q)	0.1 0.2 0.3
p (Percentile Proportion)	0.95
λ (Effect Size Xo)	1 1.2 1.4

Annotated Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Numeric Results							
Solve For:		Sample Size (N)					
Assurance Requirement:		$\Pr(Q \leq \omega) \geq 1 - \gamma$					
Interval Type:		Two-Sided Confidence Interval					
Number of Covariates:		1					
Confidence Level 1 - α	Sample Size N	Percentile Proportion p	Upper Bound of Relative Precision ω	Lower Bound of Assurance Probability		Effect Size Xo λ	Confidence Interval Width Factor CIW
				Target 1 - γ_T	Actual 1 - γ_A		
0.95	501	0.95	0.1	0.9	0.9041	1.0	0.329
0.95	568	0.95	0.1	0.9	0.9056	1.2	0.329
0.95	646	0.95	0.1	0.9	0.9057	1.4	0.329
0.95	134	0.95	0.2	0.9	0.9112	1.0	0.659
0.95	151	0.95	0.2	0.9	0.9003	1.2	0.659
0.95	172	0.95	0.2	0.9	0.9045	1.4	0.659
0.95	64	0.95	0.3	0.9	0.9054	1.0	0.991
0.95	73	0.95	0.3	0.9	0.9198	1.2	0.990
0.95	82	0.95	0.3	0.9	0.9073	1.4	0.990

Confidence Intervals for Regression-Based Reference Limits using Assurance Probability

Report Definitions

Q is the relative precision. $Q = C / R$ where C is the width of the exact confidence interval of the reference limit (percentile) and R is the width of the reference range.

$1 - \alpha$ is the confidence level of the confidence interval of the percentile.

N is the sample size of the study.

p is the percentile proportion. It is the proportion of observations that fall at or below the 100pth percentile value.

For example, a value of 0.7 indicates the 70th percentile.

ω is the upper bound of the relative precision, Q. The sample size guarantees (assures) that $100(1 - \gamma)\%$ of Q's will be less than this value.

$1 - \gamma_T$ is the target (planned) lower bound of the assurance probability.

$1 - \gamma_A$ is the actual lower bound of the assurance probability achieved by this sample size. It may be different from the target value because of the discrete nature of N.

λ is the effect size given the particular value of X_0 . If $K = 1$, $\lambda = |\mu_x - X_0|/\sigma_x$.

CIW is a factor that when multiplied by the standard deviation of the residuals (σ , not σ_x) gives the width of the confidence interval of the reference limit using ω as the value of Q.

References

Shieh, Gwonen. 2018. 'The appropriateness of Bland-Altman's approximate confidence intervals for limits of agreement.' BMC Medical Research Methodology. 18,45,1. <https://doi.org/10.1186/s12874-018-0505-y>

Hahn, G. J. and Meeker, W.Q. 1991. Statistical Intervals. John Wiley & Sons. New York.

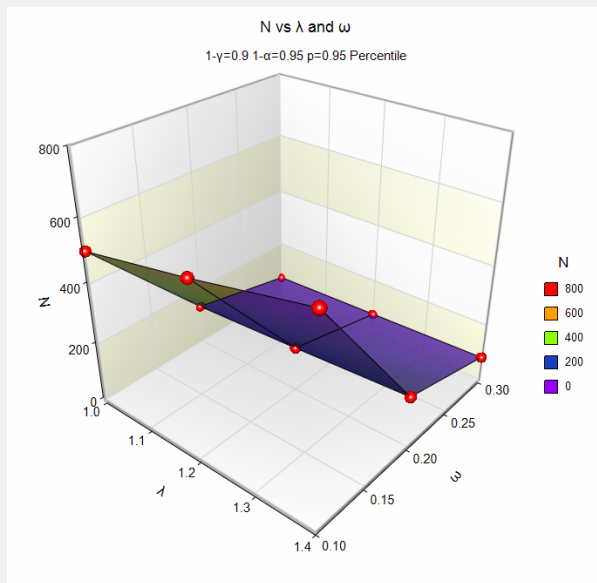
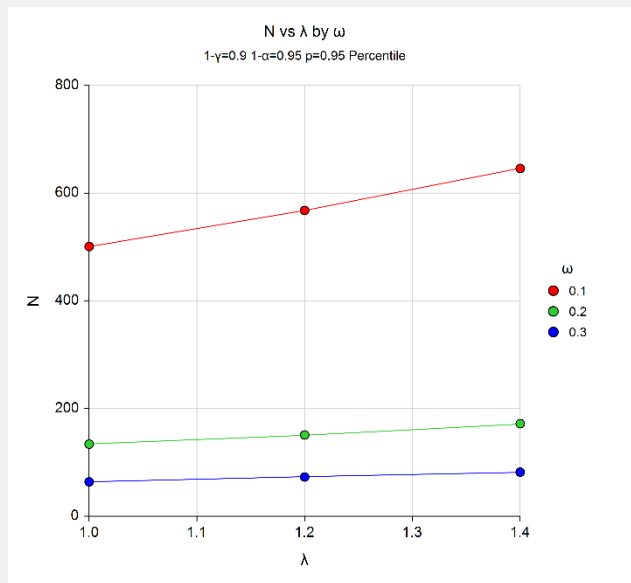
Summary Statements

A sample size of 501 produces a two-sided 95% confidence interval of the 95th percentile with a relative precision of no more than $\omega = 0.1$ with a lower bound on the assurance probability of 0.9. The effect size, λ , is 1.

This report shows the calculated sample size for each of the scenarios.

Chart Section

Chart Section



These plots show the sample sizes required for the various scenarios.

Example 2 – Validation using Shieh (2017)

Shieh (2017) page 196 gives several examples of calculating the necessary sample size for of an exact confidence interval for the 95th percentile when the confidence coefficient is 95%, λ is 1, $1 - \gamma$ is 0.90, and ω is 0.1. The number of covariates is 3. The sample size is 503.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the procedure window. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Sample Size (N)
Confidence Level ($1 - \alpha$)	0.95
K (Number of Covariates).....	3
$1 - \gamma$ (Lower Bound of Assurance Probability)	0.90
ω (Upper Bound of the Relative Precision, Q)	0.1
p (Percentile Proportion).....	0.95
λ (Effect Size Xo).....	1

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Numeric Results							
Solve For:	Sample Size (N)						
Assurance Requirement:	$\Pr(Q \leq \omega) \geq 1 - \gamma$						
Interval Type:	Two-Sided Confidence Interval						
Number of Covariates:	3						
Confidence Level	Sample Size	Percentile Proportion	Upper Bound of Relative Precision	Lower Bound of Assurance Probability		Effect Size Xo	Confidence Interval Width Factor
$1 - \alpha$	N	p	ω	Target $1 - \gamma_T$	Actual $1 - \gamma_A$	λ	CIW
0.95	503	0.95	0.1	0.9	0.9015	1	0.329

The sample size computed by **PASS** is also 503. This validates the procedure.